Longer-term (Oral, Dermal) Reference Value or Dose; Longer-term (Inhalation) Reference Value or Concentration: RfV_{LO}, RfV_{LD}, RfV_{LI}; RfD_{LO}, RfD_{LD}, RfC_{LI} or RfC_{L}

Chronic (Oral, Dermal) Reference Value or Dose; Chronic (Inhalation) Reference Value or Concentration: RfV_{CO}, RfV_{CD}, RfV_{CI}; RfD_{CO}, RfD_{CD}, RfC_{CI} or RfC_{C}

The Panel recommends that endpoint- or life stage-specific reference values such as the RfD_{DT} (reference dose for developmental toxicity), which were originally proposed in *Guidelines for Developmental Toxicity Risk Assessment* (U.S. EPA, 1991), not be derived. Rather, a sample reference value should be calculated for each relevant and appropriate endpoint and these should then be considered in the derivation of various duration reference values. Reference values should be derived to be protective of all types of effects for a given duration of exposure and are intended to protect the population as a whole, including potentially susceptible subgroups. Thus, the RfD_{DT} concept of a critical window of exposure for some health effects is addressed in the adoption of the less-than-chronic reference values. This recommendation does not preclude, however, using specific common endpoints in the assessment of cumulative risk for mixtures or chemicals that have a common mode of action or for risk management purposes.

4.3. CHARACTERIZATION OF THE EXTENT OF THE HEALTH-RELATED DATABASE FOR SETTING REFERENCE VALUES

A necessary first step in hazard characterization is the critical evaluation of all pertinent and relevant human and animal data that are available in the open literature as well as data submitted to the Agency in response to various regulatory standards, data call-ins, or other requirements and agreements.

4.3.1. Review of Studies

Data will be available from a wide variety of sources, including studies conducted according to EPA guidelines, studies conducted by industry using Organization for Economic Cooperation and Development or other protocols, experimental studies conducted by academic researchers, epidemiology studies, case reports or series, and controlled clinical studies in volunteers.[7] These studies will be of widely differing quality; EPA must evaluate each study to determine whether it is of acceptable quality.

4.3.1.1. *Adequacy of Studies*

The following list of questions could be helpful in the process of evaluating data from animal and human studies.

All types of studies:

- What was the purpose of the study and is there a clearly delineated hypothesis?
- Is there sufficient description of the protocol, statistical analyses, and results to make an evaluation?
- Were the appropriate endpoints assessed in the study?[8] Were the techniques used for the assessment scientifically sound?
- Were appropriate statistical techniques applied for each endpoint? Was the power of the study adequate to detect effects?
- Did the study establish dose-response relationships? Was a BMD lower confidence level (BMDL), LOAEL or NOAEL established?
- Is the shape of the dose-response curve consistent with the known toxicokinetics of the test compound?

---

[7]Currently, OPP is reviewing its policy concerning use of human data from studies in which there is intentional pesticide exposure, and it has asked the National Academy of Sciences for input on the acceptability of such studies and ethical criteria for their use under the Protection of Human Subjects Rule (the "Common Rule") (EPA, 2001c).

[8]A chemical may cause a variety of toxic effects depending on the amount, duration, timing, and pattern of exposure (i.e., continuous, periodic, or intermittent). These effects may range from severe—such as death—to more subtle biochemical, physiological, or pathological changes in one or more organ systems. In addition, the effects will vary depending on their latency following exposure and when the observations are made. Primary attention is given in risk assessment to those effects in the lower exposure range and/or the effects most biologically appropriate for a human health risk assessment.

- Do effects fit with what is known about mode of action?

- Is the dose-response curve for precursor events consistent with the dose-response curve for clinical effects?

- Are the results of the study biologically plausible?

- What uncertainties exist? Do the results of the study indicate the need for follow-up studies to reduce uncertainties?

- Are the study conclusions supported by the data?

Human studies:

- What were the data sources for exposure, health status, and risk factors (e.g., questionnaires, biological measurements, exposure/work history record reviews, or exposure/disease registries) and what were their strengths and limitations?

- What methods were used to control, measure, or reduce various forms of error (e.g., misclassification or interviewer bias, confounding factors and potential effect modifiers) and their potential impact on the findings? What is the validity (accuracy) and reliability (reproducibility) of the methods used to determine exposure and outcome? What were the response rates?

- What major demographic and other personal factors were examined (e.g., age, sex, ethnic group, socioeconomic status, smoking status, and occupational exposure)? What other climate or life stage factors were important for the endpoints and exposures assessed?

- Were the findings examined for biologic plausibility, internal and external consistency of the findings, and the influence of limitations of the design, data sources, and analytic methods?

Animal studies:

- Was the study sufficiently documented (e.g., conducted in accordance with good laboratory practices)?

- Were appropriate analytical techniques used to measure the stability, homogeneity, and actual level of the test substance in the study (in the water, feed, air, etc.)?

- Was an appropriate animal species used?[9] Was an appropriate number of animals used? Were sex and age considered?

- Were the dose levels appropriate? What was the basis for choosing the dose levels?

- Was an appropriate method used to assign the animals to dose groups?

- Was an appropriate route and matrix of exposure employed?[10]

- Was the duration of exposure adequate for the particular study design?

- Were possible alterations in metabolism considered at the higher exposure levels?

---

[9]The laboratory animals used most often are the rat, mouse, rabbit, guinea pig, hamster, dog, or monkey. When reviewing these studies, the risk assessor makes judgments about the ability of the study to predict the potential for toxicity in humans and tries to select data from the species that is most relevant to humans using the most defensible biological rationale. When available, comparative toxicokinetics can be used to support this decision. Absent a clearly most-relevant species, the most sensitive mammalian species is used, that is, the species that shows toxicity at the lowest exposure level.

[10]The most appropriate route of exposure is the route for which an evaluation is to be made. The toxicity of the chemical may differ with route of exposure because of differences in mechanism of action or toxicokinetics (absorption, distribution, metabolism, and excretion). Development of data to establish dosimetry for the purpose of route-to-route extrapolation is encouraged; however, route-to-route extrapolation is inappropriate when based exclusively upon default assumptions regarding exposure and toxicokinetics. Even within the same route of exposure, responses may differ due to alterations in toxicokinetics, for example, dietary or water exposure versus oral gavage.

Professional judgment is required to decide, on the basis of a thorough review of all available data and studies, whether any observed effect is adverse and how the results fit with what is known about the underlying mode of action. These judgments require the input of experts trained in toxicology, statistics, and epidemiology and, often, of specialists in the structure and function of the target organ systems. Both the biological and the statistical significance of the effects are considered when making these judgments. Biological significance is the determination that the observed effect (a biochemical change, a functional impairment, or a pathological lesion) is likely to impair the performance or reduce the ability of an individual to function or to respond to additional challenge from the agent. Biological significance is also attributed to effects that are consistent with steps in a known mode of action. Statistical significance quantifies the likelihood that the observed effect is not due to chance alone. Precedence is given to biological significance, and a statistically significant change that lacks biological significance is not considered an adverse response.

For many discrete or quantal endpoints (e.g., birth defects, tumors, or some discrete pathological changes), this judgment is more straightforward because criteria have been established for deciding what type and incidence of effects are to be considered to be adverse, and an increase above the background rate can be judged using statistical tools. In the case of continuous measures (e.g., body weight, enzyme changes, physiological measures), this tends to be more difficult, because the amount of change to be considered adverse has not been defined by toxicologists or health scientists. Consequently, the endpoint is often decided in the context of the endpoint itself, the study, and the relationship of changes in that endpoint to other effects of the agent.

Decisions about the amount of change to consider adverse must always be made using professional judgment and must be viewed in light of all the data available on the endpoint of concern. All toxicological data on a chemical must be reviewed before deciding whether an effect is biologically significant and adverse. Using a default cutoff value to define adversity for continuous measures may result in an inappropriate interpretation of data and less than optimum evaluation of a chemical's effects.

4.3.2. Issues to be Considered in Characterizing the Database for Risk Assessment

4.3.2.1. *The Weight-of-Evidence Approach*

A weight-of-evidence approach such as that provided in EPA's RfC Methodology (U.S. EPA, 1994) or in EPA's proposed guidelines for carcinogen risk assessment (U.S. EPA, 1999a) should be used in assessing the database for an agent. This approach requires a critical evaluation of the entire body of available data for consistency and biological plausibility. Potentially relevant studies should be judged for quality and studies of high quality given much more weight than those of lower quality. When both epidemiological and experimental data are available, similarity of effects between humans and animals is given more weight. If the mechanism or mode of action is well characterized, this information is used in the interpretation of observed effects in either human or animal studies. Weight of evidence is not to be interpreted as simply tallying the number of positive and negative studies, nor does it imply an averaging of the doses or exposures identified in individual studies that may be suitable as PODs for risk assessment. The study or studies used for the POD are identified by an informed and expert evaluation of all the available evidence.

4.3.2.2. *Use of Human and Animal Data in Risk Assessment*

Adequate human data are the most relevant for assessing risks to humans. When sufficient human data are available to describe the exposure-response relationship for an adverse outcome(s) that is judged to be the most sensitive effect(s), reference values should be based on human data. Much more data on a wide range of endpoints typically are required to establish confidence that there are no effects of exposure. If sufficient human data are not available to provide the basis for reference values, data from animal studies must be employed. It is advantageous if some human data are available to compare with effects observed in animals, even if the human data are not adequate for quantitative analysis. Availability of data on effects in humans at least allows qualitative comparison with effects observed in animals for determining whether toxicity occurs in the same organ systems and whether the nature of the effects is similar or different. If no human data are available, reliance must be exclusively on animal data. In that case, attention should be paid to whether data are available in more than one species and, if so, whether the same or similar effects occur in different species and possible sources of any observed differences.

One of the major default assumptions in EPA's risk assessment guidelines is that animal data are relevant for humans (e.g., U.S. EPA, 1991, 1996, 1998c). Such defaults are intended to be used in the absence of experimental data that can provide direct information on the relevance of animal data.

Several types of information should be considered when determining the relevance or nonrelevance of effects observed in animal models for humans. This information is used in a variety of ways, from determining the role of metabolism in toxicity (Is the parent chemical or a metabolite responsible for toxicity?), to assessing whether homologous activity would be expected across species (Do humans share the sensitivity of the animal model, or is the response due to some species-specific idiosyncratic reaction?), to determining whether or not a threshold is likely to exist for the response (Are repair mechanisms capable of maintaining a homeostatic process?). All of this information must be weighed in light of the known heterogeneity of the human population versus the relatively inbred status of laboratory animals used in toxicity testing studies and housed under carefully controlled environmental conditions.

Table 4-1 presents several factors to consider when evaluating the weight of evidence about the likelihood of the occurrence of effects in humans that is based on animal data (in conjunction with human data, if available). The table is not necessarily intended to delineate all factors that may need to be considered, but rather to provide a framework for evaluation and interpretation. It is important to evaluate the database in a holistic manner, determining strengths and weaknesses that are relevant to the overall assessment. Each chemical and database presents a unique set of issues that must be evaluated critically and thoughtfully.

The dose-response nature of the data is an important characteristic of the database or individual study. When data are **dose related**, that is, when the incidence and/or intensity of response changes in an orderly manner as a function of dose, the effect should be considered to be of greater importance than when there is no apparent association between exposure and toxicity. Note, however, that the dose-response relationship need not be monotonic. U-shaped (or inverted U-shaped) dose-response functions are not uncommon in toxicology. For example, a chemical may induce an enzyme at low doses and inhibit it at high doses. Similarly, many solvent-like chemicals (including alcohol) produce increased motor activity at lower doses and depressed activity at high doses.

Similarly, comparative **toxicokinetic/metabolism** data that suggest qualitative and quantitative comparability to that in humans would support the relevancy of animal data. Evidence suggesting a difference in toxicokinetics/metabolism would require additional exploration regarding whether the difference(s) results in a major qualitative or quantitative difference in internal dose in humans.

The **similarity of effects** between species is also an important aspect in characterizing the database. Similar effects in more than one species indicate that the effect provides increased weight of evidence for the risk assessment process, even if such data are not available in humans. In contrast, response data that show inconsistency of effects among studies and/or species that cannot be explained by differences in toxicokinetics/metabolism or timing and/or magnitude of exposure, may suggest that less emphasis be placed on the effect. “Similarity” does not necessarily require identical effects between species. For example, changes in motor activity in

Table 4-1. Factors for evaluation of the weight of evidence regarding the likelihood of effects in humans

Factor	Increased weight	Decreased weight
Dose-response relationship	Orderly change in effect as a function of exposure (need not be monotonic)	No identified relationship between exposure and magnitude of effect
Toxicokinetics/ metabolism	Qualitative and quantitative comparability between humans and animals	Qualitative and quantitative differences between humans and animals
Similarity of effects	Similar effects in more than one animal species or in animals and humans	Inconsistency of effects among studies and/or species that cannot be explained by differences in timing and/or magnitude of exposure or toxicokinetics/metabolism
Mode of action	Demonstration of homologous mode of action in animal model and humans	Evidence suggesting that the mode of action is species specific and irrelevant to humans
Temporal relationship	Consistent temporal relationship between exposure and effect	Lack of temporality between exposure and effect

animals evaluated in the neurotoxicity screening test and cognitive effects in humans would generally be considered similar, because both are indicative of changes in nervous system function.

Mode of action information is also important in understanding whether a particular effect may be important for humans. For example, a transient reduction in anogenital distance in the postnatal animal following perinatal exposure to an anti-androgen has increased weight if the chemical is also known to act as an anti-androgen in humans. Likewise, the interpretation of increased skeletal variants observed following exposure to many chemicals would be enhanced by data indicating that the mechanistic pathways for these agents and the overall biological significance defined were also a possibility in humans. Mode of action data are also important in determining whether various chemicals work by common modes or mechanisms of action, which would then be considered in a cumulative risk assessment.

Another criterion that is important in evaluating data is the **temporal relationship** between exposure and effect. The exposure should precede the effect at an interval that is consistent with what is known about the toxicokinetics and mode of action of the agent. It may be the case, however, that higher doses produce a shorter latency to effect than do lower doses.

4.3.2.3. *Characterization of Effects in Potentially Susceptible Subpopulations*

A dose-response analysis for potentially susceptible subpopulations should be done as part of the overall dose-response analysis for health effects in general. “Susceptible” in this context means a differential (greater) response at the same internal dose in a particular segment of the population due to intrinsic (possibly unknown) factors. “Susceptible subpopulations” is used here to refer both to life stages and to other factors that may predispose individuals to greater response to an exposure. Life stages may include the developing individual before and after birth up to maturity (e.g., embryo, fetus, young child, adolescent), adults, or aging individuals. Other susceptible subpopulations may include people with specific genetic polymorphisms that render them more vulnerable to a specific agent or people with specific diseases or pre-existing conditions (e.g., asthmatics). The term may also refer to gender differences, lifestyle choices, or nutritional state.

It is important to recognize that little basis currently exists for a priori identification of susceptible subpopulations for many chemicals. Without other data to raise suspicions, only the evaluation of effects in various segments of the population such as those mentioned above can identify susceptible subpopulations for a particular chemical and a particular set of exposure conditions. In some situations, differential exposure rather than differential susceptibility per se may be the critical issue (e.g., hand-to-mouth activity in toddlers). Economic differences may also result in differential exposure and susceptibility.

A great deal of attention has been given in recent years to the issue of children as a susceptible subpopulation. Several approaches have been proposed for characterizing the database concerning the potential pre- and postnatal toxicity of a particular chemical and providing some guidance as to the weight of evidence or degree of concern for children’s health. However, each approach has been developed for a slightly different purpose and, as such, is generally complementary to, but not the same as, the other approaches.

EPA’s developmental toxicity (U.S. EPA, 1991) and reproductive toxicity (U.S. EPA, 1996) risk assessment guidelines describe an approach that characterizes the database as sufficient or insufficient to judge whether a chemical does or does not pose a hazard within the context of dose, route, duration, and timing of exposure. The International Programme on Chemical Safety (IPCS) (IPCS, 1995) proposed an approach based on the quality of information gathered in developmental and reproductive toxicity studies and the types of data that were not available from these studies. EPA’s draft 10X toxicology report (U.S. EPA, 1999b) further extended the recommendations for characterizing risks to children’s health within the context of the FQPA by discussing issues that would increase or decrease the level of concern.

The present report endorses and extends the recommendations of the 10X Toxicology Working Group's report by incorporating the issues dealing with level of concern into a framework for evaluating the evidence regarding the identification and characterization of susceptible subpopulations (see below). A workshop was held recently to discuss aspects of a framework for children's health risk assessment and to emphasize a broader perspective on the issues that should be considered in hazard characterization, dose-response assessment, exposure assessment, and risk characterization for children as a susceptible subpopulation (ILSI RSI, 2001).

In contrast with the attention paid to children and asthmatics as potentially susceptible subpopulations in recent years, little attention has been focused on risk assessment for other potentially susceptible subgroups. As outlined in Chapter 3, there currently are no requirements in EPA animal study protocols for exposure during old age or for outcome evaluations near the end of the life span following earlier life stage exposures. Similarly, healthy animals that are more genetically homogeneous than humans are used in standard toxicity testing protocols, and information on pre-existing conditions or genetic polymorphisms is largely unavailable from animal studies.

Human studies also usually employ healthy nonelderly individuals, although some studies in more susceptible populations have been conducted, such as studies of the effects of air pollutants in asthmatics. Individuals who have identified risk factors that are not the focus of a study are usually excluded from the study sample. It is important to consider such characteristics of the database if human data are used as the basis for the risk assessment.

As can be seen in Table 4-2, several issues must be considered in assessing the potential for some subpopulations, including different life stages, to have greater susceptibility than others to a chemical. These include the **timing (life stage)-response relationship,** indicating greater susceptibility to exposure at some life stages than at others; whether effects are of a **different type** in identifiable subgroups of the population; and the **dose-response relationship**, that is, whether effects are observed at different levels of exposure in different subpopulations.

Another important consideration is whether effects are observed at the same dose but with a shorter **latency** in different subpopulations. Additionally, differences among groups in terms of the **seriousness** and **reversibility of effects** must be considered. For example, an agent may produce relatively mild and reversible neurological effects in adults but produce permanent behavioral impairment following in utero exposure. It is also important to keep in mind that effects that may initially appear to be reversible may re-appear later or be predictive of later adverse outcomes. This is probably best exemplified by certain outcomes following a

Table 4-2. Factors for evaluating evidence regarding identification and characterization of susceptible subpopulations[a]

Factor	Increased weight	Decreased weight
Timing (life stage) - response relationship	Effects occur at greater magnitude at one or more life stage(s)	No difference in effects at different life stage(s)
Type of effect	Different types of effects in specific subpopulations	Same effect(s) across all potential subpopulations
Dose-response relationship	Effect occurs at lower exposures in one or more subpopulation(s)	No evidence for differential dose-response across different subpopulations
Latency of effect	Latency to observed effect different in specific subpopulations	No difference between subpopulations in latency to effect
Seriousness/ reversibility of effects	Effects different in seriousness or degree of reversibility in specific subpopulations and/or differences in later consequence of an initially reversible effect	No differences between subpopulations in seriousness and/or reversibility of effects, or in later consequences of an initially reversible effect

[a] Subpopulations may be defined by gender, individuals at different life stages (fetus, child, adult, elderly), differences in genetic polymorphisms, and/or pre-existing diseases or conditions that may result in differential sensitivity to adverse effects from exposure to a specific toxic agent.

developmental exposure; for example, an initial depression in birth weight or weight gain or subtle developmental retardation may be indicators of more serious abnormalities later in life.

4.3.3. Characterization of the Extent of the Database

The derivation of an RfD or an RfC is a multifaceted process that involves the coordination of data gathering and evaluation, analysis and judgment in varying proportions, and integration of all the information available. A vital part of the chronic RfD and RfC derivation process that relies heavily on judgment, for example, is the current approach to characterizing the database. For example, the minimum dataset for low-confidence and high-confidence RfDs and RfCs has been specifically defined as follows (U.S. EPA, 1994, 2002c): *minimum dataset for a low confidence chronic RfD or RfC* is a single subchronic study. The *minimum dataset for* *a high confidence chronic RfD or RfC* is a chronic study in two species, a single two-generation reproductive toxicity study, and a developmental toxicity study in two species by the appropriate route of exposure.

The Technical Panel is recommending a somewhat different approach. Instead of specifying particular studies, this approach emphasizes the types of data needed (in terms of both human and animal data) for deriving reference values and recommends the use of a narrative description of the extent of the database rather than a single confidence statement. The Technical Panel believes that this approach encourages the use of a wider range of information in deriving reference values that take into consideration the issues of duration and route of exposure, the timing of exposures, the types and extent of endpoint assessment (i.e., structural and function), the susceptible subpopulations evaluated, and the potential for latent effects and/or reversibility of effects. In addition, this approach encourages the identification of data that would be needed or useful for improving the risk assessment for a particular chemical or group of chemicals.

To characterize the database, the Technical Panel has developed a description of a "minimal" database and a "robust" database as a way of describing the range of data that can be used for deriving a reference value (Box 4-3). A great deal of scientific judgment is necessary when evaluating the extent of the database for a particular chemical. Defining the extent of the database requires an overall evaluation and judgment as to where in the minimal–robust continuum the available database should be characterized. The Technical Panel purposely did not define additional categories between minimal and robust (moderate), and the Panel has serious concerns about developing such categories because of the tendency to try to characterize a database with single word descriptors. Instead, we strongly support a narrative description of the extent of the database, with emphasis on the strengths and limitations of the data. It should also be noted that a database that is less than minimal should not be used to derive a reference value.

Rather than presenting separate "minimal" and "robust" database descriptions for each type of reference value that might be derived, the descriptions in Box 4-3 are intended to apply generally across the various reference value types (e.g., acute, short-term, longer-term , or chronic durations for oral, dermal, or inhalation routes of exposure). Additionally, it is expected that the different types of reference values for a particular chemical will be developed within the same assessment. In this manner, the entire database for a chemical may be relied upon in the development of each of the different values (e.g., important and relevant insights may be gleaned from toxicity studies for exposure durations other than those directly corresponding to the type of reference value being developed).

A minimal database as defined above can be used to set reference values, but the limitations of such a database should be clearly recognized and discussed in the narrative description. For example, a minimal database may provide data on only one duration or route of exposure or it may be specific to only one endpoint or organ system. Thus, the uncertainties related to such a database will be great and should be reflected in the size of the UFs applied for reference value derivation (see further discussion below).

Box 4-3. Description of minimal and robust databases

Minimal Database: no human data available, route-specific toxicity data are limited to dose-response data applicable to the duration in question with assessment of endpoints other than mortality. A study showing only effect levels for mortality or other extremely severe toxicity would not be sufficient to set a reference value.

Robust Database: includes extensive human and/or animal toxicology data that cover route-specific information on many health endpoints, durations of exposure, timing of exposure, life stages and susceptible subpopulations. In the absence of complete human data, mechanistic and other data show the relevance of the animal data for predicting human response. Specifically, the dose-response data for the reference value in question includes endpoint-specific data (e.g., developmental toxicity, neurotoxicity) coupled with toxicokinetic information as needed for route-to-route extrapolation. The toxicity studies include the evaluation of a variety of endpoints (e.g., hematological, clinical, histology of target organs) and endpoints specific to any known hazard characterization. The database for a reference value of less-than-chronic duration has also addressed the issue of reversibility of effects and latency to response, taking into consideration the possibility that less-than-chronic exposure may lead to effects at some period of time after exposure. Biological and chemical characteristics of the exposure and outcomes, as well as known limits on reserve capacities and repair of damage, form the basis for determining the appropriate length of follow-up.

On the other hand, a robust database would address issues of potential toxicity in humans and animals and include data on several durations and routes of exposure as well as a thorough assessment of a variety of health endpoints. It would also include sufficient data on toxicokinetics and mode action to provide extensive information for extrapolation of effects to humans, including potentially susceptible subpopulations. A complete database on a single health endpoint that does not contain information on other endpoints of possible relevancy would not necessarily constitute a robust database, nor would a database that provides complete information on one route and/or duration of exposure be considered robust.

It is clear that a robust database represents a “gold standard” that will rarely, if ever, be available. However, a lack of robustness does not mean that the database is deficient to the extent that a reference value could not be derived or that large UFs would need to be applied. Sound scientific judgement will be required to determine which UFs are appropriate in each case.

A critical assessment of the extent and quality of the database will inform the selection of the endpoints to be used to derive the reference values and the appropriate UFs. A reference value based on a single study would likely have a high degree of uncertainty. As more information from additional toxicology studies, toxicokinetic studies, structure-activity relationships, and human data becomes available, EPA can have greater assurance that the appropriate species, route of exposure, and target organ system(s) are known for each duration reference value needed for a human health risk assessment. As this additional information becomes available, the use of UFs will likely decrease. The ultimate objective is to account for all human health endpoints resulting from exposures over all life stages from before conception to the elderly adult.

The optimum assessment considers subtle effects that impact an individual's quality of life as well as so-called "frank" effects (death and major disease). The evaluation should encompass immediate health outcomes as well delayed responses to an exposure (i.e., latent responses), although most current testing guidelines do not explicitly evaluate latency to response.

4.3.3.1. *Extent of the Database*

The following series of questions regarding the extent of the database can help guide the assessment process:

- Have adequate studies been conducted to establish the target organs/endpoints?

- Have the effects been characterized for both sexes and all life stages?

- Are data pertaining to potentially susceptible subpopulations available?

- Are the responses consistent across species? Are the results of the studies biologically plausible?

- Is the route and matrix of exposure relevant to the specific reference value being derived?

- Is the duration of exposure appropriate for the specific reference value being derived?

- Is the animal species and strain appropriate for extrapolation to humans?

- To what degree may the biological endpoints be extrapolated (qualitatively and quantitatively) to humans?

- Are toxicokinetic data available? Are they available for both sexes, for relevant life stages, for other susceptible subpopulations?

- Is the shape of the dose-response curve consistent with the known toxicokinetics of the test compound?

- Are the metabolism and toxicokinetics in the animal species similar to those of humans?

- Has the dose-response curve been replicated by or is it consistent with data from other laboratories and other test species?

- Have the data for all relevant endpoints been adequately modeled by the BMD or other appropriate quantitative analysis to determine the most sensitive endpoint(s)?

- How well is the toxicity characterized? Do the results of all the studies indicate the possibility of effects on particular systems that have not yet been explored sufficiently or do they indicate that additional studies may reveal effects not yet characterized?

4.4. DERIVATION OF REFERENCE VALUES

After the database has been thoroughly evaluated for quality and extent, as outlined above, several decisions must be made and procedures applied before the final derivation of a reference value. This section summarizes the current procedures and points out assumptions made and areas for improvement and clarification. A variety of factors related to the derivation of reference values is discussed, including the selection of relevant endpoints for the POD for various duration reference values (Section 4.4.1). Adjustment of the study dose/exposure for duration is described in Section 4.4.2, and derivation of a HED or HEC is discussed in Section 4.4.3.